Art Driven by Visual Representations of Chemical Space

Daniela Gaytán-Hernández,[1] Ana L. Chávez-Hernández,[1] Edgar López-López,[1,2] Jazmín Miranda-Salas,[1] Fernanda I. Saldívar-González,[1] José L. Medina-Franco*[,1]

[1] *DIFACQUIM Research Group, Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de México, Avenida Universidad 3000, México City 04510, Mexico*

[2] *Department of Chemistry and Graduate Program in Pharmacology, Center for Research and Advanced Studies of the National Polytechnic Institute, Mexico City 07000, Mexico*

*Contact author: medinajl@unam.mx

Abstract: Science and art have been connected for centuries. With the development of new computational methods, new scientific disciplines have emerged, such as computational chemistry, and related fields, such as chemoinformatics that use informatic methods to solve chemical problems focusing on small molecules. Chemoinformatics is grounded on the chemical space concept: a multi-descriptor space in which chemical structures are described. In several practical applications, visual representations of the chemical space of compound datasets are low-dimensional plots useful for identifying patterns. However, the authors propose that the plots can also be used as artistic expressions. This manuscript introduces an approach to merging art with chemoinformatics through visual and artistic representations of chemical space. As case studies, we portray the chemical space of food chemicals and other chemicals to generate visually appealing graphs with twofold benefits: sharing chemical knowledge and developing pieces of art driven by chemoinformatics. The art driven by chemical space visualization will help increase the application of chemistry and art and contribute to general education and dissemination of chemoinformatics and chemistry through artistic expressions.

Keywords: artwork; chemical space; chemical multiverse; chemoinformatics; data visualization; education; food chemistry; foodinformatics; molecular representation; open science.

Abbreviations: AI, artificial intelligence; ECPF, extended connectivity fingerprint; HBD, hydrogen bond donors; HBA, hydrogen bond acceptors; GTM, generative topographic mapping; LogP, partition coefficient octanol/water; MACCS, Molecular ACCes System; MW, molecular weight; PCA, principal component analysis; TMAP, TreeMap; t-SNE, t-distributed stochastic neighbor embedding; TPSA, topological polar surface area; SOM, self-organizing map.

1. Introduction

Art can be considered as the set of activities and products of human beings with aesthetic, ethical, and communication objectives that impact individuals or societies[1]. Its impact may seek to transmit ideas, emotions,

needs, concerns, or values[2]. Science can be considered as an art´s tool: the one that makes the materialization of ideas possible and delimits the ideas of artists. What is important about science is not only that it has served to enable the work to be executed. What is fundamental is that it has allowed it to be imagined. Furthermore, scientific knowledge allows for a more profound interpretation of works of art.

Historically, the relationship between science and art has existed since humans created art. One example is chemistry, a scientific discipline that historically has had a symbiotic relationship with art and has determined its respective evolutions. Among the many interactions of chemistry in art are: the development of pigments and spectroscopic techniques, materials for conservation and restoration, to name just a few[3,4].

The advent of computers gave rise first to computational chemistry, and then, chemoinformatics. Chemoinformatics, also frequently referred to in the literature as cheminformatics[5] aims to manage and organize information, visualize chemical space, perform data mining, and establish mathematical relationships between chemical structures and properties. While bioinformatics focuses on biologically relevant macromolecules, chemoinformatics is focused on small compounds[6]. As an independent theoretical discipline, chemoinformatics relies on the chemical space concept[7–10]. Understanding the concept of chemical space within and outside chemoinformatics can be complicated. Generally, this concept has been accompanied by various images that seek to represent characteristics that chemists have assigned according to the inherent purposes of their research, leaving aside the aesthetic composition that, in turn, can contribute to deepening and communicating beyond the common sense, which associates thinking to an operation that excludes its connections with the affections, sensitivity, and creation. In Chemoinformatics, chemical space has been defined as a chemical descriptor vector space (cf. Figure 1A) set by the numerical vector X encoding property or molecular structure aspects as elements of the descriptor vector X[11]. As such, chemoinformatics methods strongly depend on molecular representation and numerical descriptors[12]. There are many descriptors whose selection will depend on the type of molecules studied, for example, organic, inorganic, small molecules, peptides (whose size can differ significantly), natural products, and food chemicals, to name a few. For small molecules (e.g., molecular weight < 1000 Da), it is common to use as descriptors molecular fingerprints[13,14], whole molecule properties (e.g., properties of pharmaceutical relevance[15,16]), and sub-structures such as molecular scaffolds[17]. Figure 1A shows a schematic representation of the concept of chemical space e.g., chemical space table as a matrix where compounds are the rows and the numerical descriptors are the

columns. Graphical techniques and reduction dimension techniques are used to map the usually large multi-dimensional spaces into two or three dimensions that can be plotted and easily visualized.

Since the chemical space of a set of compounds is not unique and will depend on the set of descriptors chosen to describe it, there are possible multiple chemical spaces that are theoretically possible for the same data set. Continuing this line of thinking, a chemical multiverse was proposed recently and defined as "the group of numerical vectors that describe differently the same set of molecules." An alternative definition of the chemical multiverse is a "group of multiple chemical spaces, each one defined by a given set of descriptors - a group of "descriptor universes"[7]. The chemical multiverse concept is represented in Figure 1B.

Chemical spaces and chemical multiverses are, like many other types of analysis, frequently analyzed by means of data visualization techniques (Figure 1). Indeed, data visualization is widely used in science and other areas to effectively summarize and communicate data to produce information and, ultimately, knowledge. Extensive reviews have been published concerning the visualization of chemical spaces[9,10]. As reviewed, there are multiple methods of visualization, such as principal component analysis (PCA)[18], t-distributed stochastic neighbor embedding (t-SNE)[19], Tree MAP (TMAP)[20], self-organizing map (SOM)[21–23], and the generative topographic mapping (GTM)[24]. Each one will have advantages and disadvantages. As emphasized above, the visualization of a given data set will depend on the type of descriptors used.

The visual representation of chemical spaces can lead to visually appealing figures, particularly if appropriate color schemes are used. The visually attractive settings are used to emphasize patterns in the chemistry data to facilitate visual information extraction. For instance, to highlight grouping or clustering in the chemistry data or to rapidly identify patterns in the structure-property landscapes. At the same time, the visually attractive graphs can be for the chemistry expert and non-expert, a visually attractive graph or sort of a digital "painting" or work of art. In other words, the graph or digital painting is driven by chemical structures and descriptors. Therefore, the person generating the chemical space representation could be considered a chemical space artist that can communicate not only chemical data and information but even emotions if the chemical structures are associated with a personal, emotional, or another type of feeling the "artist" / author want to communicate through the visualization, e.g., an artistic expression.

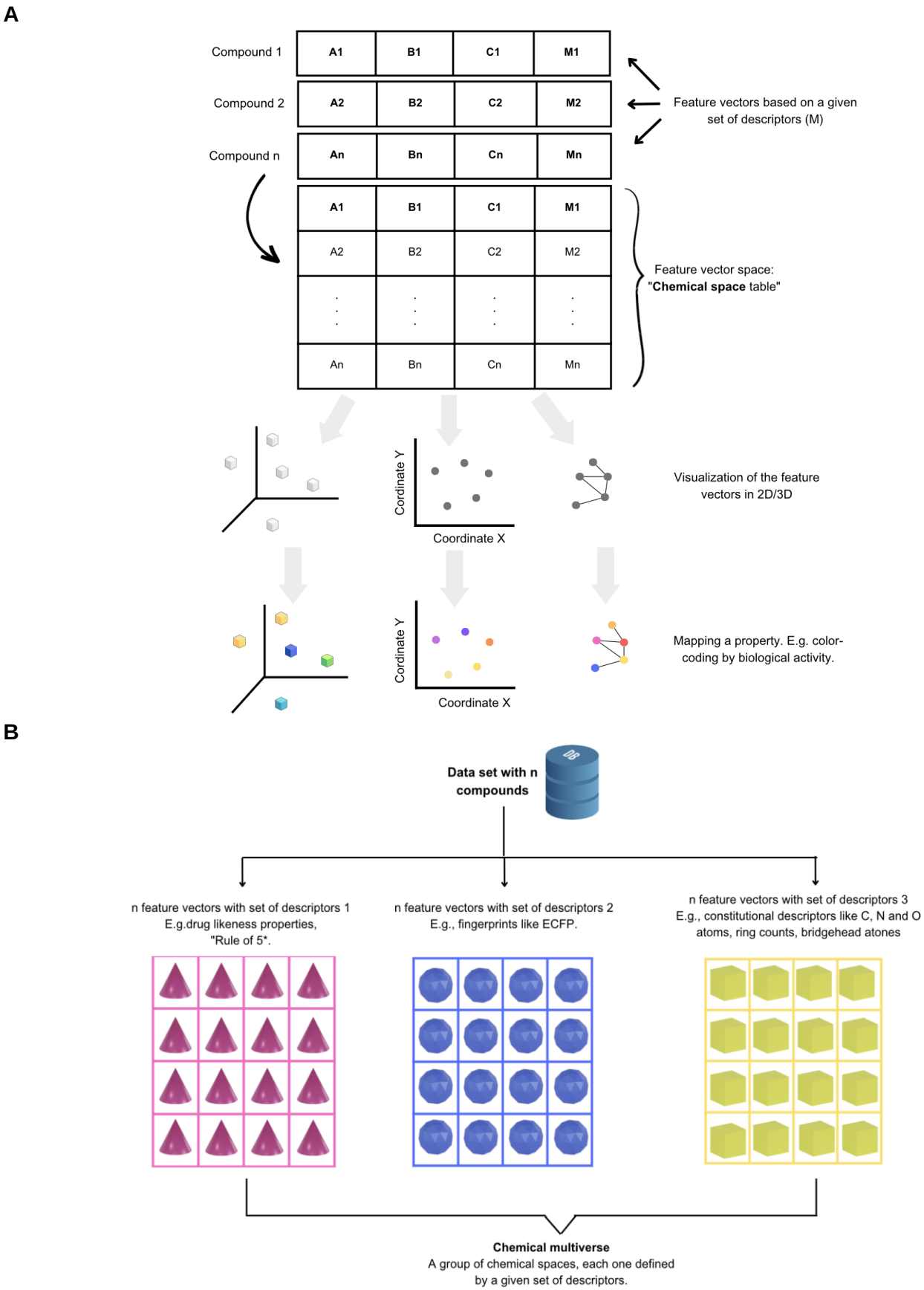


Figure 1. Schematic concept of **A**) chemical space and its visual representation in low-dimensions. **B**) Schematic representation of a chemical multiverse for a hypothetical data set of n compounds: descriptors of different design (continuous properties, molecular fingerprints, constitutional descriptors, etc.) can lead to alternative chemical spaces for the same data set.

In this sense, the concept of chemical space also opens up the possibility of searching for new representations that have to do with the need to configure another image of thought, and think in a novel fashion; it is a creative task and is similar to art.

This manuscript aims to propose the general notion of generating visual representations of chemical space and chemical multiverses as a means of chemical communication that produce new experiences and, in parallel, artistic expressions. To illustrate the proposal, we generated chemical space visualizations of four flavor categories from a large public database of food chemicals, FooDB[25], using different descriptors and molecular fingerprints. For the examples, we considered four flavor categories, as detailed in the Methods section. The concept would further promote art driven by chemoinformatics and can be expanded to other information-related disciplines, such as bioinformatics. Using different descriptors and visualization methods, we show examples of chemical multiverse visualizations of four flavor categories from FooDB and other chemical compounds.

2. Methodos

2.1. Data sets

Herein, we used food chemicals to generate visual representations of the chemical space as artworks. Food and its flavors, colors, textures and aromas are generally associated with the great pleasures of life and for this reason, they have been a source of inspiration in the world of art. However, an approximation at the structural level of the molecules has not been addressed. Specifically, we used chemical structures from the public database FooDB[25]. The current version of FooDB contains 70,477 compounds, and after data set standardization (described in detail in section 2.2) has 52,856 molecules. FooDB has information about macronutrients and micronutrients, and food chemicals that give food flavor, color, taste, texture, and aroma to foods. Each chemical item in FooDB contains more than 100 separate data fields providing detailed compositional, biochemical, and physiological information[25]. From FooDB, 4,964 natural flavorings derived from food compounds were identified across twenty flavor categories. Figure 2 summarizes the frequency of the seven most populated categories.

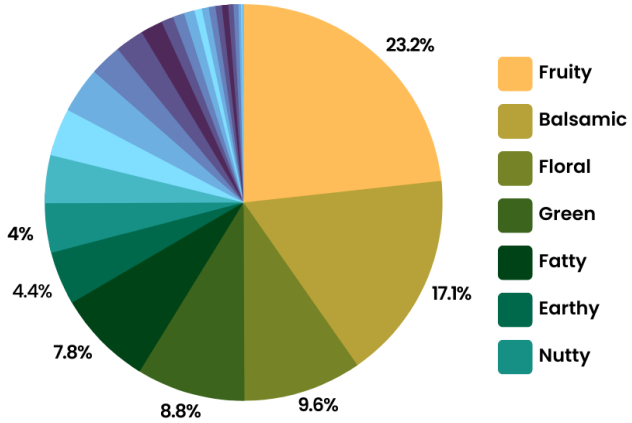


Figure 2. The seven most frequent flavor categories identified in FooDB.

From the twenty-seven flavor categories, we defined four new flavor categories: 1) ground flavors, 2) wine-tasting, 3) contrast between spicy and fatty, and 4) natural remedies. Table S1 in the Supplementary material shows the number of compounds in each of the four categories considered in this work. Flavors of the ground/flavor similar to herbaceous are composed of earthy, herbaceous, and green flavors. Wine-tasting are composed of fruity and floral flavors. The contrast between spicy and fatty is composed of spicy and fatty flavors. Medicinal is composed of balsamic, chemical, and medicinal, which are characteristic flavors found in ointments, alcohol, and syrups. Figure S2 in the Supplementary material shows the number of overlapping compounds between the selected flavor categories.

2.2. Data set standardization

Compounds in FooDB, encoded as SMILES strings[12], were standardized using the open-source cheminformatics toolkit RDKit[26] and Standardizer, LargestFragmentChoser, Uncharger, Reionizer y TautomerCanonicalizer functions implemented in MolVS[27]. Compounds with valence errors or any chemical element other than H, B, C, N, O, F, Si, P, S, Cl, Se, Br, and I were removed. Stereochemistry information, when available, was retained. Compounds with multiple components were split, and the largest component was retained. The remaining compounds were neutralized and reionized to generate the corresponding canonical tautomer.

2.3. Molecular descriptors

For each molecule, physicochemical properties, and molecular fingerprints were calculated as descriptors using Python language and RDKit. The whole molecule descriptors computed were hydrogen bond donors

(HBD), hydrogen bond acceptors (HBA), topological polar surface area (TPSA), number of rotatable bonds (RB), molecular weight (MW), and partition coefficient octanol/water (LogP). Molecular fingerprints computed were Molecugrounds System (MACCS) Keys (166-bits)[13], extended connectivity fingerprint (ECFP)[14] of 1024-bits with diameter 4 (ECFP4). Of note, virtually any other descriptors can be used, as further commented in the Discussion, Section 4.

2.4. Visualization methods

In this study, we used three well-known dimensionality reduction methods: t-SNE, PCA, and TMAPs although additional visualization methods can be used. Briefly, t-SNE generates plots that organize compounds. Similar compounds form clusters, and dissimilar compounds are distant from each other. PCA is a linear dimensionality reduction technique to transform data with many dimensions (i.e., descriptors) into a lower dimensional space and keep the different relationships between the data points as much as possible[18]. PCA was generated from six whole molecule descriptors (MW, HB, HBA, SlogP, TPSA, and RB). TMAPs allow visualizing many chemical compounds through the distance between clusters and the detailed structure of these through branches and sub-branches. Local sensitive hashing allows each compound to be grouped hierarchically according to common substructures using molecular fingerprints. In this work, we use MACCS keys (166-bits)[13] fingerprint. Then, each chemical compound was encoded using the MinHash algorithm. The number of nearest neighbors, k = 50, and the factor used by the augmented query algorithm, kc = 10, were used to generate the TMAPs[20].

3. Results

Figures 3-6 show examples of so-called “Art Galleries” composed by visualization of the chemical space of different food chemical categories. The visual representations of chemical space were generated with t-SNE (Figures 3 and 4), PCA (Figure 5), and TMAPs (Figure 6). Below each image (i.e., “digital paintings”) is presented basic information of the “technique” (visualization method, allusive to the techniques used in paintings), descriptors, and chemicals (that would be meaningful information for a chemistry-oriented person to understand the data presented). Each visual representation of the chemical space or Artwork includes a “Title” that is reminiscent of the name of the piece of art or digital painting.

A

Title: ***Mindful taste I***
Technique: t-SNE
Descriptors: MW, HB, HBA, SlogP, TPSA, and RB.
Chemicals: Flavor food-chemical compounds.
Flavor food category: Green, Earthy, and Herbaceous.

B

Title: *Mindful taste II*
Technique: t-SNE
Descriptors: MW, HB, HBA, SlogP, TPSA, and RB.
Chemicals: Flavor food-chemical compounds.
Flavor food category: Fruity and Floral.

C

TITLE: *Mindful taste III*
Technique: t-SNE
Descriptors: MW, HB, HBA, SlogP, TPSA, and RB.
Chemicals: Flavor food-chemical compounds.
Flavor food category: Fatty and Spicy.

D

TITLE: *Mindful taste IV*
Technique: t-SNE
Descriptors: MW, HB, HBA, SlogP, TPSA, and RB.
Chemicals: Flavor food-chemical compounds.
Flavor food category: Balsamic, Chemical, and Medicinal

Figure 3. Four flavor categories and full FooDB. The flavor categories are **A**) Ground flavors (655 compounds), **B**) Wine-tasting (1024 compounds), **C**) Contrast between spicy and fatty (430 compounds), and **D**) Natural remedies (762 compounds).

A

TITLE: *Umami*
Technique: t-SNE
Descriptors: MW, HB, HBA, SlogP, TPSA, and RB.
Chemicals: Flavor food-chemical compounds.
Flavor food category: Green, Earthy, and Herbaceous.

B

TITLE: *Sweet*
Technique: t-SNE
Descriptors: MW, HB, HBA, SlogP, TPSA, and RB.
Chemicals: Flavor food-chemical compounds.
Flavor food category: Fruity and Floral.

C

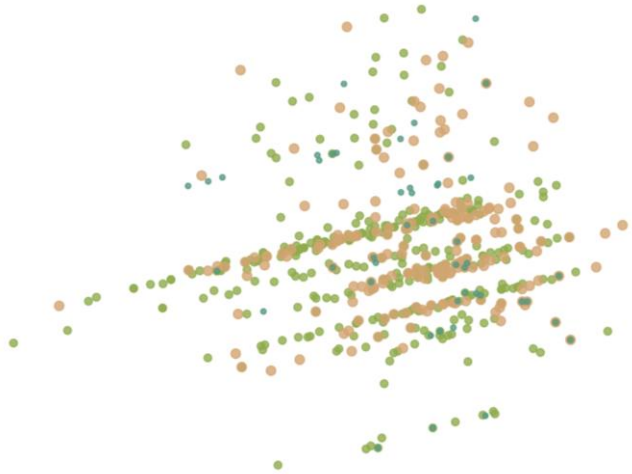
TITLE: *Sour*
Technique: t-SNE
Descriptors: MW, HB, HBA, SlogP, TPSA, and RB.
Chemicals: Flavor food-chemical compounds.
Flavor categories: Fatty and Spicy.

D

TITLE: *Bitter*
Technique: t-SNE
Descriptors: MW, HB, HBA, SlogP, TPSA, and RB.
Chemicals: Flavor food-chemical compounds.
Flavor categories: Balsamic, Chemical, and Medicinal.

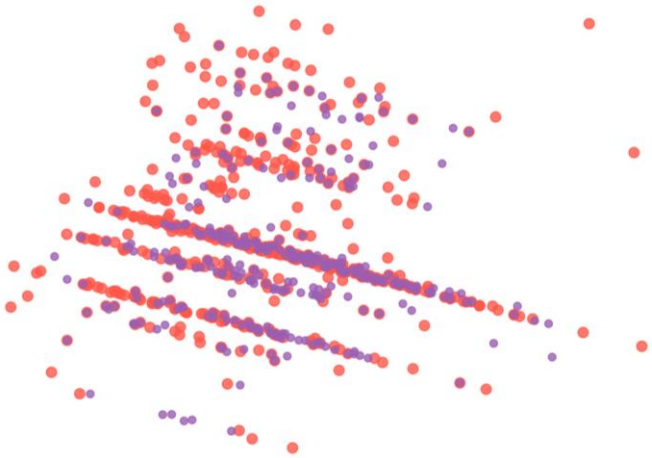
Figure 4. Four flavor categories: **A**) Ground flavors (655 compounds), **B**) Wine-tasting (1024 compounds), **C**) Contrast between spicy and fatty (430 compounds), and **D**) Natural remedies (762 compounds).

A



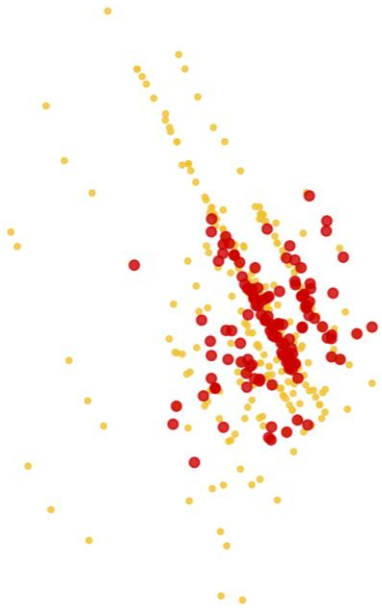
TITLE: *PANTE*
Technique: PCA
Descriptors: MW, HB, HBA, SlogP, TPSA, and RB.
Chemicals: Flavor food-chemical compounds.
Flavor categories: Green, Earthy, and Herbaceous.

B



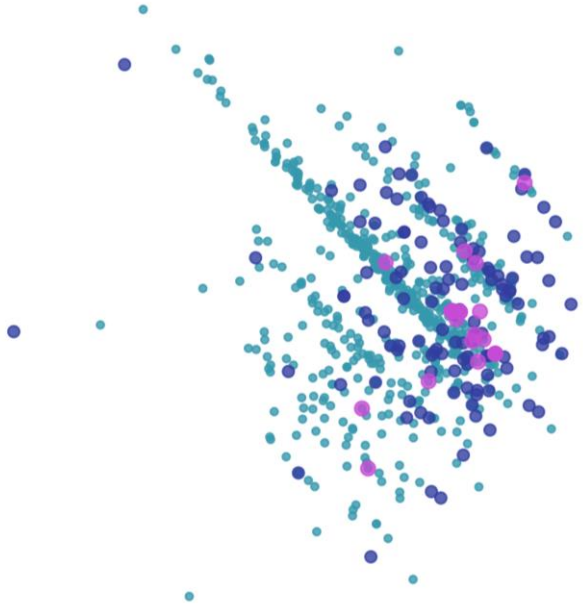
TITLE: *SYRAH*
Technique: PCA
Descriptors: MW, HB, HBA, SlogP, TPSA, and RB.
Chemicals: Flavor food-chemical compounds.
Flavor categories: Fruity and Floral.

C



TITLE: *FONTINA*
Technique: PCA
Descriptors: MW, HB, HBA, SlogP, TPSA, and RB.
Chemicals: Flavor food-chemical compounds.
Flavor categories: Fatty and spicy.

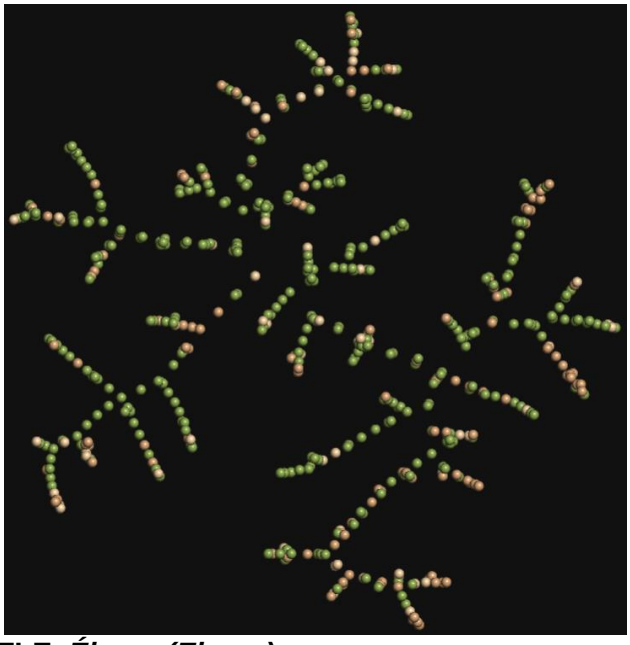
D



TITLE: *XOCOC*
Technique: PCA
Descriptors: MW, HB, HBA, SlogP, TPSA, and RB.
Chemicals: Flavor food-chemical compounds.
Flavor categories: Balsamic, Chemical, and Medicinal.

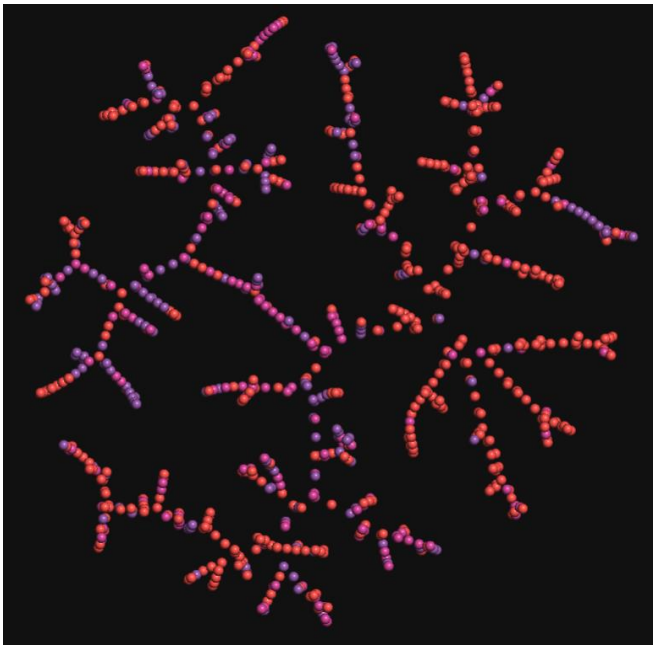
Figure 5. Four flavor categories: **A**) Ground flavors (655 compounds), **B**) Wine-tasting (1024 compounds), **C**) Contrast between spicy and fatty (430 compounds), and **D**) Natural remedies (762 compounds).

A



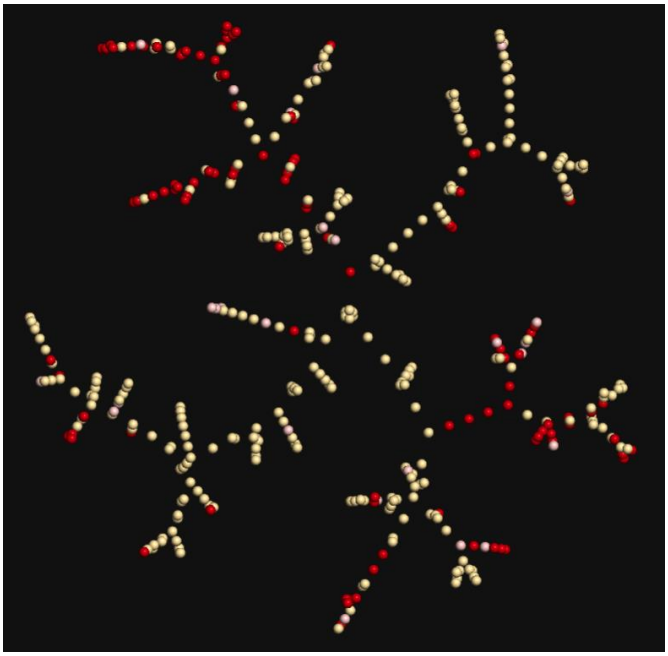
TITLE: ***Ébano (Ebony)***
Technique: TMAP
Descriptors: MACCS keys fingerprints
Chemicals: Flavor food-chemical compounds.
Flavor categories: Green, Earthy, and Herbaceous.

B



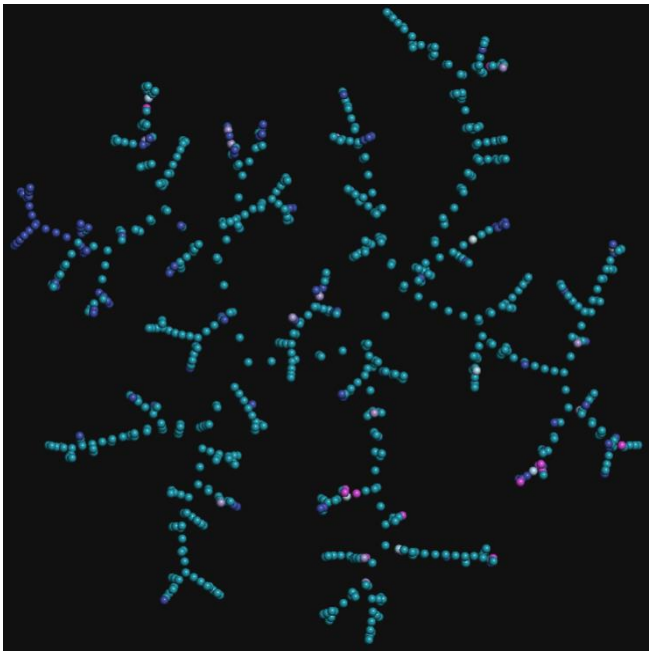
TITLE: ***Flor de corazón (Heart flower)***
Technique: TMAP
Descriptors: MACCS keys fingerprints
Chemicals: Flavor food-chemical compounds.
Flavor categories: Fruity and Floral

C



TITLE: ***Amaranto (Amaranth)***
Technique: TMAP
Descriptors: MACCS keys fingerprints
Chemicals: Flavor food-chemical compounds.
Flavor categories: Fatty and Spicy

D



TITLE: Huele a miel (*Honey´s smell*)
Technique: TMAP
Descriptors: MACCS keys fingerprint
Flavor food-chemical compounds.
Flavor categories: Balsamic, Chemical, and Medicinal

Figure 6. Four flavor categories: **A**) Ground flavors (655 compounds), **B**) wine-tasting (1024 compounds), **C**) Contrast between spicy and fatty (430 compounds), and **D**) Natural remedies (762 compounds).

4. Discussion

Chemoinformatics has been broadly used in drug discovery. Still, it has many more applications in chemistry, with increasing applications in food chemistry, as evidenced by the emergence of the research areas of food chemical informatics or food informatics[28,29]. There are others, such as natural products [30,31], polymers, and materials, to name a few[6]. Herein, we propose expanding the realm of chemoinformatics´ applications through the visual representation of the chemical space of compound data sets - herein illustrated with food chemicals - to yield exemplary “art pieces.” The connection or synergy between chemoinformatics and art has a strong potential to bring together at least two sectors of the population that might be otherwise disconnected. From an educational point of view, which is a central need in chemoinformatics - the synergy might attract young students and kids to chemistry through art.

The subdiscipline of food informatics was proposed in 2014 as a specific application of chemoinformatics to food chemistry[28]. Since then, numerous applications of chemoinformatics to different aspects of food chemistry have been published, including analysis of the chemical space of food chemicals to characterize the structural diversity[32]. In section 3 we showed examples of visual representations of the chemical space of food chemicals as an artistic expression and scientific dissemination through art. There are many possibilities to expand the genesis of the proposed “art-cheminformatics,” as further elaborated in Section 5.

4.1. Exemplary art-related chemical spaces and multiverses

The examples of visual representation of chemical space as artistic representations presented in Section 3 are focused on food chemicals and molecular descriptors suitable to represent such chemical compounds. Also, examples of visualization methods used in the previous section are t-SNE, PCA, and TMAPs. However, as commented in the Introduction, the number of established visualization techniques, molecular descriptors, and, perhaps most importantly, the number of chemical structures are immense. Therefore, there are literally thousands or millions of ways to generate chemical space-driven works of art. To glimpse the artistic possibilities, Table 1 summarizes examples of the cheminformatics-driven visualization of chemical space and multiverses. The table summarizes examples of compound data sets with chemicals of different types that could be used to represent their vastness, complexity, diversity, and chaotic intrinsic features from an artistic perspective. Many more compound data sets and multiple combinations of descriptors and visualization

techniques could be used. However, as with any other artistic vehicle, the real importance of any type of art is its capacity to tell histories or convey a message that sometimes is hidden.

Table 1. Exemplary potential paintings based on the visualization of the chemical space of compound data sets.

Data set	Artistic meaning	Artwork name
Random compounds	Aleatory molecules represent the vastness of our universe and daily life. We are in contact with many chemicals every time, but we don't look at their complexity and intrinsic disorder in our universe and daily life.	"Chaos"
Diverse data set	The diversity offers many colors, flavors, tastes, and experiences. In nature, diversity (in all senses) is a constant feature.	"Diversity"
Marine natural products	We don't understand the sea; It has life, death, color, and darkness. It's constantly changing.	"The Ocean" "Immensity"
Drugs approved for the treatment of HIV	Everything happens in a positive HIV human; Fear, memories, happiness, and normality. The drugs help... but are not a complete answer.	"Living with AIDS"
Hormones - neurotransmitters.	Love = hormones + neurotransmitters + special persons.	"The chemistry of love"
Chemicals associated with depression	Depression = hormones + neurotransmitters - purpose	"Darkness"
Food chemicals	The great pleasures of life are often accompanied by flavors, colors, textures, and aromas.	"Bellyful" "Flavor trip"
ZINC database vs. drug-like compounds	We know a lot about our nature and composition, but we don't know much more. Our knowledge is a mere stain on an entire canvas that we do not yet understand.	"Our knowledge"

To illustrate further the potential of generating artistic representations through visualization of chemical space, Figure 7 shows an example of chemical space artwork from a random natural products dataset, decoding by their side effects descriptors (e.g., mutagenesis, tumorogenesis, and negative reproductive effects, etc.). Their color palette, from red to blue, represents the probability of each natural product generating side effects. The "canvas" was "painted" with a dotted technique which reflects another possible set of textures that can be developed with this type of technique. Like in Figure 7, we intrinsically know that "nature" is not always healthy and that within us, there is a delicate balance that is very easy to break.

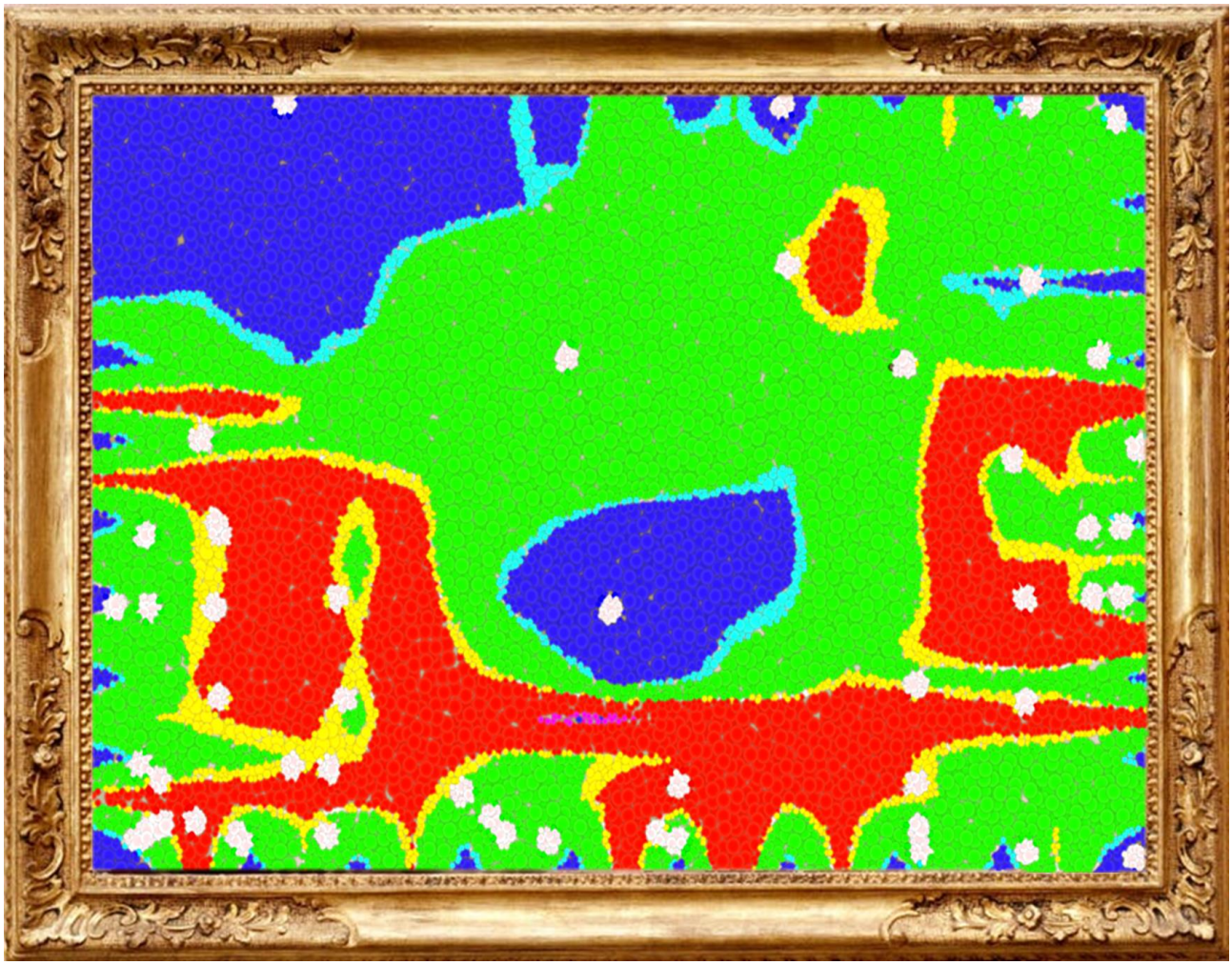
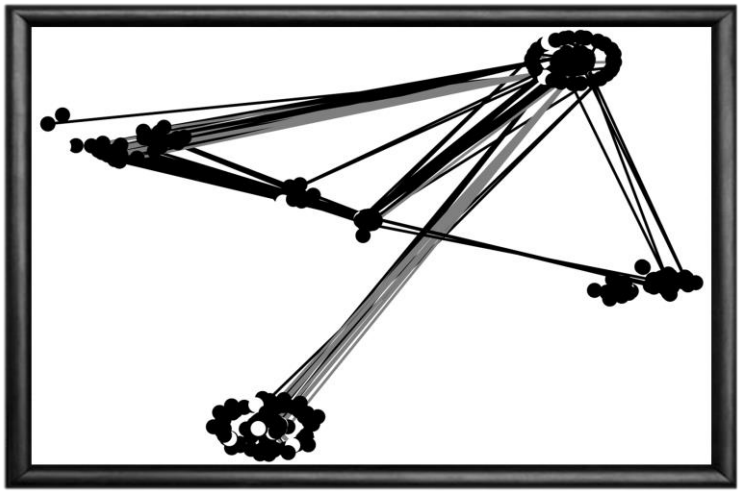
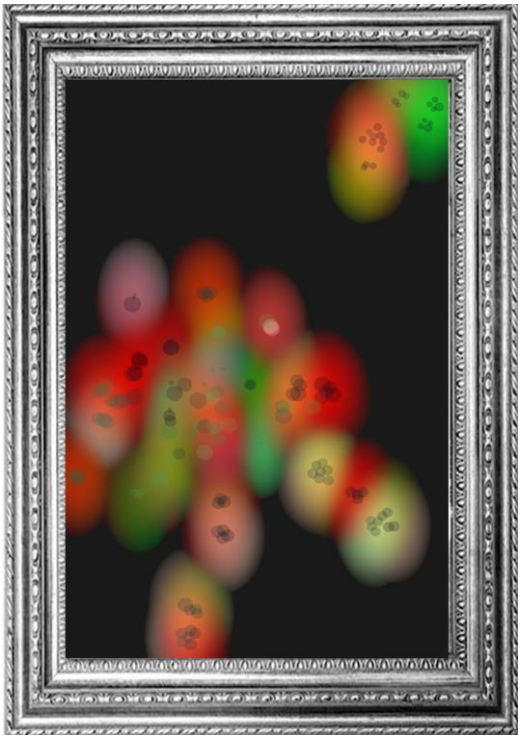


Figure 7. Chemical space art example. **Title**: “Wise nature”; **Autor**: Edgar López-López; **Technique**: SOM - using DataWarrior software[33,34]; **Dataset**: Random natural products (1000 compounds); **Descriptors**: predicted mutagenic, tumorogenic, Reproductive effective, and Irritant; **Technical description**: Each white point is a natural product, the regions colored in red represent the chemical space with a high predicted probability of containing compounds witch side effects, the opposite for the blue color; **Artistic interpretation**: The "nature" is not always healthy, in nature, there has always been a duality between what fills us with life and what takes it away.

Figure 8 shows additional examples of chemical space artwork that combine different reduction data methods and descriptors to generate an artistic visual representation of the chemical data. We encourage the readers to reflect and find other artistic interpretations that these figures could have.



TITLE: Chemical umbrella
Technique: PCA + Data fusion (chemical multiverse approach).
Descriptors: Cell-based and enzymatic inhibition data. Dots are connected based on their inhibitory activity against different types of cytochromes (proteins related to hepatic protection).
Hepatotoxic compounds.



TITLE: Broked cancer
Technique: Constellation plots.
Descriptors: Anticancer cell inhibition data.
Anticancer drugs.

Figure 8. Chemical space art examples. Chemical artworks were generated with public data [35–37].

4.2. Artificial intelligence and digital art

Artificial intelligence (AI) is used to generate artistic representations [38,39]. Although it is not the central point of this manuscript, Figure 9 illustrates images generated with free resources using keywords associated with "chemical space". Specifically, the figure shows an example of a chemical multiverse/chemical space driven by an AI-web server training on words. Although the images are attractive, a striking difference with the chemical space artworks presented in previous sections (Figures 3-8) is that the images in Figure 9 are based on keywords training, and the former are derived directly from chemical structures encoded with molecular

descriptors Another important aspect is that in the former representations there is a greater understanding and human intervention, something questionable in AI-guided representations.

A

B

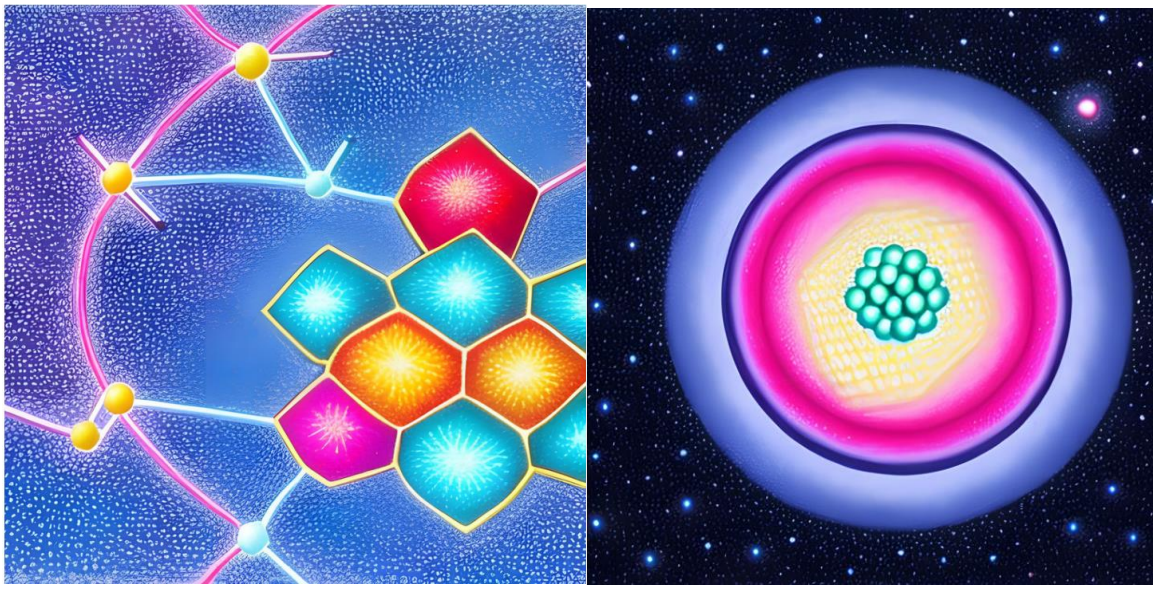


Figure 9. Example(s) of artificial intelligence-driven art with the free application Canva (https://www.canva.com/) using the keyword chemical space and **A**) Watercolor and **B**) color pencil.

5. Conclusions and outlook

Science and art have long been intimately related. A typical example is summarized by the phrase, “Drug discovery is as much an art as it is a science.” Certainly, chemistry is substantially used in art, such as in art restoration and preservation. But there is an emerging trend to apply chemistry and its concepts to generate artwork. Herein, we discuss an approach to combining art with chemoinformatics through the visual

representations of chemical space. We presented a few examples of chemical space artworks that can be “digital paintings.” The author of the low-dimensional graphs can use the plots with dual general purposes: communicate data and generate chemical information (as normally done with the visualizations of chemical space) and communicate an emotional or personal meaning to the graph (driven by chemistry and informatics principles). We also conclude that chemical space-driven works of art can be tools to promote science in general and chemistry in particular for the broad audience. The concept of chemical space seen from an artistic approach can help the general public to see molecules beyond individual entities and common representations. Thus, chemistry informatic-driven artistic expressions can be an approach to disseminating science. Such an approach aligns with the graphical abstracts frequently used in peer-reviewed journals. The "chemical art" could be useful to represent complex data but using an artistic and attractive perspective. The person generating the chemical space representation could be considered a “chemical space artist.”

We envision several further developments and areas of opportunity for art driven by visual representations of chemical space. Table 2 summarizes ongoing chemical arts projects, from the generation of “easy to use” tools, the first chemical art gallery, and the implementation of this artistic mode to introduce the new generation of chemoinformaticians to the chemical space concept. In parallel, AI methods will continue expanding and exploring the chemical space, offering new types of molecules and descriptors that could be used to increase the possibilities of representing chemical space from an artist's perspective.

Table 2. Representative developments of combining art with chemoinformatics through artistic visualizations of chemical space.

Development	Putative outcome or application
Continue developing a digital collection focused on the artistic representation of the chemical space.	*The Chemical Space Art Museum*
Generate automated workflows using open software or informatic tools to improve the accessibility of this kind of art to people with different academic/artistic backgrounds.	*ChemArt Generators*
Establish a free, open-access, and permanent repository of art pieces. This encourages open science and open art. The scientific and artistic community could support the repository.	*ChemART Gallery*
Set up a sustained educational or cultural program as a continued open and permanent exposition.	*Art Driven by Chemical Space Visualization* program.

Supplementary material

Table S1. The number of flavor compounds, flavor notes, and flavor categories. **Figure S2**. Unique and overlapping structures of four flavor categories from FooDB.

Acknowledgments

We are grateful for the rich and useful discussions with Dra. Karina Martínez-Mayorga. D.G.-H. thanks the support from CONACyT. A.L.-C.-H., E.L.-L., and F.I.S.-G thank CONACyT, Mexico, for the Ph.D. scholarships number 847870, 894234, and 848061, respectively.

References

(1) Galván-Madrid, J. L. La Química y el Arte: ¿Cómo mantener el vínculo? *Educ. Quím.* **2011**, *22* (3), 207–211, DOI: 10.1016/S0187-893X(18)30136-8.

(2) Bello, D. G. *La química de lo bello*, 2nd ed.; Ediciones Paidós, 2023.

(3) Orna, M. V. Chemistry, Color, and Art. *J. Chem. Educ.* **2001,** *78* (10), 1305, DOI: 10.1021/ed078p1305.

(4) Kafetzopoulos, C.; Spyrellis, N.; Lymperopoulou-Karaliota, A. The Chemistry of Art and the Art of Chemistry. *J. Chem. Educ.* **2006**, *83* (10), 1484, DOI: 10.1021/ed083p1484.

(5) Miranda-Salas, J.; Peña-Varas, C.; Valenzuela Martínez, I.; Olmedo, D. A.; Zamora, W. J.; Chávez-Fumagalli, M. A.; Azevedo, D. Q.; Castilho, R. O.; Maltarollo, V. G.; Ramírez, D.; Medina-Franco, J. L. Trends and Challenges in Chemoinformatics Research in Latin America. *Artif. Intell. Life Sci.* **2023**, *3* (1), 100077, DOI: 10.1016/j.ailsci.2023.100077.

(6) López-López, E.; Bajorath, J.; Medina-Franco, J. L. Informatics for Chemistry, Biology, and Biomedical Sciences. *J. Chem. Inf. Model.* **2020**, *61* (1), 26-35, DOI: 10.1021/acs.jcim.0c01301.

(7) Medina-Franco, J. L.; Chávez-Hernández, A. L.; López-López, E.; Saldívar-González, F. I. Chemical Multiverse: An Expanded View of Chemical Space. *Mol. Inf.* **2022**, *41* (11), 2200116, DOI: 10.1002/minf.202200116.

(8) Medina-Franco, J. L.; Sánchez-Cruz, N.; López-López, E.; Díaz-Eufracio, B. I. Progress on Open Chemoinformatic Tools for Expanding and Exploring the Chemical Space. *J. Comput.-Aided Mol. Des.* **2022**, *36* (5), 341–354, DOI: 10.1007/s10822-021-00399-1.

(9) Osolodkin, D. I.; Radchenko, E. V.; Orlov, A. A.; Voronkov, A. E.; Palyulin, V. A.; Zefirov, N. S. Progress in Visual Representations of Chemical Space. *Expert Opin. Drug Discov.* **2015**, *10* (9), 959-973, DOI: 10.1517/17460441.2015.1060216.

(10) Medina-Franco, J.; Martinez-Mayorga, K.; Giulianotti, M.; Houghten, R.; Pinilla, C. Visualization of the Chemical Space in Drug Discovery. *Curr. Comput. Aided Drug Des.* **2008**, *4* (4), 322–333, DOI: 10.2174/157340908786786010.

(11) Saldívar-González, F. I.; Medina-Franco, J. L. Approaches for Enhancing the Analysis of Chemical Space for Drug Discovery. *Expert Opin. Drug Discov.* **2022**, *17* (7), 789-798, DOI: 10.1080/17460441.2022.2084608.

(12) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28* (1), 31-36, DOI: 10.1021/ci00057a005.

(13) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL Keys for Use in Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (6), 1273–1280, DOI: 10.1021/ci010132r.

(14) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50* (5), 742-754, DOI: 10.1021/ci100050t.

(15) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **2001**, *46* (1-3), 3–26, DOI: 10.1016/s0169-409x(00)00129-0.

(16) Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.* **2002**, *45* (12), 2615-2623, DOI: 10.1021/jm020017n.

(17) Singh, N.; Guha, R.; Giulianotti, M. A.; Pinilla, C.; Houghten, R. A.; Medina-Franco, J. L. Chemoinformatic Analysis of Combinatorial Libraries, Drugs, Natural Products, and Molecular Libraries Small Molecule Repository. *J. Chem. Inf. Model.* **2009**, *49* (4), 1010-1024, DOI: 10.1021/ci800426u.

(18) Greener, J. G.; Kandathil, S. M.; Moffat, L.; Jones, D. T. A Guide to Machine Learning for Biologists. *Nat. Rev. Mol. Cell Biol.* **2021**, *23* (1), 40–55, DOI: 10.1038/s41580-021-00407-0.

(19) van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf?fbcl (accessed Jun 1,

2023).

(20) Probst, D.; Reymond, J.-L. Visualization of Very Large High-Dimensional Data Sets as Minimum Spanning Trees. *J. Cheminf.* **2020**, *12* (1), 1-13, DOI: 10.1186/s13321-020-0416-x .

(21) Kohonen, T. *Self-Organizing Maps*; Springer Berlin Heidelberg, 2001; pp 105-176.

(22) Schneider, P.; Tanrikulu, Y.; Schneider, G. Self-Organizing Maps in Drug Discovery: Compound Library Design, Scaffold-Hopping, Repurposing. *Curr. Med. Chem.* **2009**, *16* (3), 258-266, DOI: 10.2174/092986709787002655.

(23) Virshup, A. M.; Contreras-García, J.; Wipf, P.; Yang, W.; Beratan, D. N. Stochastic Voyages into Uncharted Chemical Space Produce a Representative Library of All Possible Drug-like Compounds. *J. Am. Chem. Soc.* **2013**, *135* (19), 7296-7303. DOI: 10.1021/ja401184g

(24) Bishop, C. M.; Svensén, M.; Williams, C. K. I. Developments of the Generative Topographic Mapping. *Neurocomputing* **1998**, *21* (1), 203-224, DOI: 10.1016/S0925-2312(98)00043-5.

(25) FooDB https://foodb.ca/ (accessed Apr 20, 2023).

(26) RDKit https://www.rdkit.org (accessed 08 January 2022).

(27) MolVS https://molvs.readthedocs.io/en/latest/ (accessed Jan 08, 2022).

(28) *Foodinformatics: Applications of Chemical Information to Food Chemistry*; Martinez-Mayorga, K., Medina-Franco, J. L., Eds.; Springer International Publishing: Cham, Switzerland, 2014.

(29) Peña-Castillo, A.; Méndez-Lucio, O.; Owen, J. R.; Martínez-Mayorga, K.; Medina-Franco, J. L. Chemoinformatics in Food Science. In *Applied Chemoinformatics*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2018; pp 501–525.

(30) Kirchmair, J. Molecular Informatics in Natural Products Research. *Mol. Inf.* **2020**, *39* (11), 2000206, DOI: 10.1002/minf.202000206.

(31) Medina-Franco, J. L.; Saldívar-González, F. I. Cheminformatics to Characterize Pharmacologically Active Natural Products. *Biomolecules* **2020**, *10* (11), 1566, DOI: 10.3390/biom10111566.

(32) Naveja, J. J.; Rico-Hidalgo, M. P.; Medina-Franco, J. L. Analysis of a Large Food Chemical Database: Chemical Space, Diversity, and Complexity. *F1000Res.* **2018**, *7*, DOI: 10.12688/f1000research.15440.2.

(33) Sander, T.; Freyss, J.; von Korff, M.; Rufener, C. DataWarrior: An Open-Source Program for Chemistry Aware Data Visualization and Analysis. *J. Chem. Inf. Model.* **2015**, *55* (2), 460-473, DOI:

10.1021/ci500588j.

(34) López-López, E.; Naveja, J. J.; Medina-Franco, J. L. DataWarrior: An Evaluation of the Open-Source Drug Discovery Tool. *Expert Opin. Drug Discov.* **2019**, *14* (4), 335-341, DOI: 10.1080/17460441.2019.1581170.

(35) López-López, E.; Medina-Franco, J. L. Towards Decoding Hepatotoxicity of Approved Drugs through Navigation of Multiverse and Consensus Chemical Spaces. *Biomolecules* **2023**, *13* (1), 176, DOI: 10.3390/biom13010176.

(36) Medina-Franco J. L.; Naveja J. J.; López-López E. Reaching for the Bright StARs in Chemical Space. *Drug Discov. Today* **2019**, *24* (11), 2162-2169, DOI: 10.1016/j.drudis.2019.09.013.

(37) López-López, E.; Cerda-García-Rojas, C. M.; Medina-Franco, J. L. Tubulin Inhibitors: A Chemoinformatic Analysis Using Cell-Based Data. *Molecules* **2021**, *26* (9), 2483, DOI: 10.3390/molecules26092483.

(38) DALL·E 2 https://openai.com/dall-e-2/ (accessed Jun 20, 2023).

(39) Ramesh, A.; Dhariwal, P.; Nichol, A.; Chu, C.; Chen, M. Hierarchical Text-Conditional Image Generation with CLIP Latents, 2022, DOI: arXiv:2204.06125.

GRAPHICAL ABSTRACT